

A new class of variance reduction techniques using lattice symmetries

Thomas Blum,^{1,2} Taku Izubuchi,^{3,2} and Eigo Shintani²

¹*Physics Department, University of Connecticut, Storrs, CT 06269-3046, USA*

²*RIKEN-BNL Research Center, Brookhaven National Laboratory, Upton, NY 11973, USA*

³*Brookhaven National Laboratory, Upton, NY 11973, USA*

We present a general class of unbiased improved estimators for physical observables in lattice gauge theory computations which significantly reduces statistical errors at modest computational cost. The error reduction techniques, referred to as covariant approximation averaging, utilize approximations which are covariant under lattice symmetry transformations. We observed cost reductions from the new method compared to the traditional one, for fixed statistical error, of 16 times for the nucleon mass at $M_\pi \sim 330$ MeV (Domain-Wall quark) and 2.6-20 times for the hadronic vacuum polarization at $M_\pi \sim 480$ MeV (Asqtad quark). These cost reductions should improve with decreasing quark mass and increasing lattice sizes.

PACS numbers: 11.15.Ha, 12.38.Gc, 07.05.Tp

As non-perturbative computations using lattice gauge theory are applied to a wider range of physically interesting observables, it is increasingly important to find numerical strategies that provide precise results. In Monte Carlo simulations our reach to important physics is still often limited by statistical uncertainties. Examples include hadronic contributions to the muon's anomalous magnetic moment [1], nucleon form factors and structure functions [2], including nucleon electric dipole moments [3–6], hadron matrix elements relevant to flavor physics (*e.g.*, $K \rightarrow \pi\pi$ amplitudes) [7], and multi-hadron state physics [8], to name only a few.

As a generalization of low-mode averaging (LMA) [9, 10], we present a class of unbiased statistical error reduction techniques, utilizing approximations that are covariant under lattice symmetry transformations. LMA has worked well in cases where low eigenmodes of the Dirac operator dominate [11]: low energy constants in the ε -regime [9, 12–15], pseudoscalar meson masses and decay constants [16–18], and so on. With a modest increase in computational cost, the generalized method can reduce statistical errors by an order of magnitude, or more, even in cases where LMA fails.

Unlike LMA, we account for all modes of the Dirac operator, averaging over (most of) the lattice volume, with modest additional computational cost. The all-to-all methods [19, 20] implement this stochastically for the higher modes, while treating the low-modes exactly. For expectation values invariant under translations, statistics effectively increase by averaging over the whole lattice. The all-to-all method is advantageous when the stochastic noise introduced in the target observable is comparable to, or smaller than, the gauge field fluctuations of the ensemble [21], which typically holds only for many random source vectors per measurement. The error reduction techniques presented here, which do not rely on stochastic noise, are potentially more effective, provided an inexpensive approximation can be found for the desired observable.

In lattice gauge theory simulations an ensemble of gauge field configurations $\{U_1, \dots, U_{N_{\text{conf}}}\}$ is generated randomly, according to the Boltzmann weight, $e^{-S[U]}$, where $S[U]$ is the lattice-regularized action. The expectation value of a primary, covariant observable, \mathcal{O} ,

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \mathcal{O}[U_i] + \mathcal{O} \left(\frac{1}{\sqrt{N_{\text{conf}}}} \right), \quad (1)$$

is estimated as the ensemble average, over a large number of configurations, $N_{\text{conf}} \sim \mathcal{O}(100 - 1000)$. Here, we primarily consider observables made of fermion propagators $S_F[U]$ computed on the background gauge configuration U .

By exploiting lattice symmetry transformations $g \in G$, that transform $U \rightarrow U^g$, a general class of variance reduction techniques is introduced. First construct an *approximation* $\mathcal{O}^{(\text{appx})}$ to \mathcal{O} which must fulfill the following conditions,

appx-1: $\mathcal{O}^{(\text{appx})}$ should fluctuate closely with \mathcal{O} ,

$$r \equiv \text{Corr}(\mathcal{O}, \mathcal{O}^{(\text{appx})}) = \frac{\langle \Delta \mathcal{O} \Delta \mathcal{O}^{(\text{appx})} \rangle}{\sqrt{\langle (\Delta \mathcal{O})^2 \rangle \langle (\Delta \mathcal{O}^{(\text{appx})})^2 \rangle}} \approx 1,$$

and $\langle (\Delta \mathcal{O})^2 \rangle \approx \langle (\Delta \mathcal{O}^{(\text{appx})})^2 \rangle$, where $\Delta X = X - \langle X \rangle$.

appx-2: the cost to compute $\mathcal{O}^{(\text{appx})}$ is smaller than \mathcal{O} 's, $\text{cost}(\mathcal{O}^{(\text{appx})}) \ll \text{cost}(\mathcal{O})$.

appx-3: $\langle \mathcal{O}^{(\text{appx})} \rangle$ is *covariant* under a lattice symmetry transformation, $g \in G$, $\langle \mathcal{O}^{(\text{appx})}[U^g] \rangle = \langle \mathcal{O}^{(\text{appx}),g}[U] \rangle$ (in the examples below, a stronger condition holds: $\mathcal{O}^{(\text{appx})}$ is covariant on each configuration, rather than on average, $\mathcal{O}^{(\text{appx})}[U^g] = \mathcal{O}^{(\text{appx}),g}[U]$).

Note $\mathcal{O}^{(\text{appx})}$ and $\mathcal{O}^{(\text{appx}),g}$ refers to the approximations before and after applying a symmetry transformation g .

Using \mathcal{O} and $\mathcal{O}^{(\text{appx})}$ one can define an improved ob-

servable

$$\begin{aligned}\mathcal{O}^{(\text{imp})} &= \mathcal{O}^{(\text{rest})} + \mathcal{O}_G^{(\text{appx})}, \\ \mathcal{O}^{(\text{rest})} &= \mathcal{O} - \mathcal{O}^{(\text{appx})}, \quad \mathcal{O}_G^{(\text{appx})} = \frac{1}{N_G} \sum_{g \in G} \mathcal{O}^{(\text{appx}),g},\end{aligned}\quad (2)$$

where an average over N_G symmetry transformations in G is taken.

For **appx-1**, the statistical error of $\langle \mathcal{O}^{(\text{imp})} \rangle$ is

$$\text{err}(\text{imp}) \approx \text{err} \sqrt{2(1-r) + \frac{1}{N_G}}, \quad (3)$$

which can be made smaller than the original (err) by a judicious choice of $\mathcal{O}^{(\text{appx})}$. The fluctuation from $\mathcal{O}^{(\text{rest})}$, the first term in (3), is suppressed due to $r \approx 1$, while the second term is reduced by $1/N_G$ without too much additional cost as required by **appx-2** (correlations among \mathcal{O} , $\mathcal{O}^{(\text{appx})}$, and $\mathcal{O}^{(\text{appx}),g}$ have been ignored, which is a good approximation for noisy observables or large volume). Due to covariance, **appx-3**, it is easy to prove the ensemble averages of (primary observables) $\mathcal{O}^{(\text{appx})}$, $\mathcal{O}^{(\text{appx}),g}$, and $\mathcal{O}_G^{(\text{appx})}$ are all equal, so the improved estimator (2) is unbiased, $\langle \mathcal{O}^{(\text{imp})} \rangle = \langle \mathcal{O} \rangle$.

The idea of exploiting covariance [9, 10] to improve statistical errors has a wider range of applicability than LMA, so in general we call it *covariant approximation averaging* (CAA). Several comments on CAA follow. From Eq. 3 the accuracy of the approximation $\mathcal{O}^{(\text{appx})} \approx \mathcal{O}$ (**appx-1**) should be precise enough so that the statistical error from $\mathcal{O}^{(\text{rest})}$ is below, say, one-half of the desired final precision. Too accurate an approximation wastes resources. In $\mathcal{O}^{(\text{imp})}$, most of the statistical fluctuation is carried by $\mathcal{O}^{(\text{appx})}$, which is reduced by averaging over $N_G (\gg 1)$ measurements with smaller cost (**appx-2**). Balance between these opposing parts of the method allows CAA to reduce statistical errors significantly while keeping the computational cost low.

In the framework of CAA the best choice of approximation depends on the target observables and lattice parameters such as quark mass and volume. In principle, any set of lattice symmetries, G , can be used in CAA. We limit ourselves to the case of translation symmetries in the following examples.

The first example is LMA. In LMA eigen-systems of the Hermitian Dirac operator are obtained for the part of the spectrum closest to zero,

$$D_H v_i = \lambda_i v_i, \quad (i = 1, 2, \dots, N_{\text{eig}}), \quad (4)$$

$$0 < |\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_{N_{\text{eig}}}| = \lambda_{\text{cut}}, \quad (5)$$

which is then used to construct, through spectral decomposition, the low-mode approximation of the fermion

TABLE I. LMA and AMA algorithms

LMA algorithm	AMA algorithm
1: Compute low-modes v_i of D_H	1: if $\lambda_{\text{cut}} \neq 0$, $N_{\text{eig}} > 0$ Compute low-mode v_i of D_H
2: Set source b and G -invariant initial guess x_0	
3: Compute exact S and $\mathcal{O}[S]$ precisely (use deflation if v_i exists)	
4: Repeat for S_{LM} in (6) and $\mathcal{O}^{(\text{appx})} = \mathcal{O}[S_{\text{LM}}]$	4: Repeat for S_{AM} in (8) and $\mathcal{O}^{(\text{appx})} = \mathcal{O}[S_{\text{AM}}]$ using deflated CG (if $\lambda_{\text{cut}} \neq 0$)
5: $\mathcal{O}^{(\text{rest})} = \mathcal{O}[S] - \mathcal{O}[S_{\text{LM}}]$	5: $\mathcal{O}^{(\text{rest})} = \mathcal{O}[S] - \mathcal{O}[S_{\text{AM}}]$
6: Set shifted source b^g and G -invariant initial guess x_0^g	
7: Average $\mathcal{O}^{(\text{appx}),g} = \mathcal{O}[S_{\text{LM}}]$ over $g \in G$ to get $\mathcal{O}_G^{(\text{appx})}$	7: Average $\mathcal{O}^{(\text{appx}),g} = \mathcal{O}[S_{\text{AM}}]$ over $g \in G$ to get $\mathcal{O}_G^{(\text{appx})}$
8: $\mathcal{O}^{(\text{imp})} = \mathcal{O}^{(\text{rest})} + \mathcal{O}_G^{(\text{appx})}$	

propagator,

$$S_{\text{LM}}(x, y) = \sum_{i=0}^{N_{\text{tot}}} v_i(x) f_{\text{LM}}(\lambda_i) v_i^\dagger(y), \quad (6)$$

$$f_{\text{LM}}(\lambda) = \frac{1}{\lambda} \theta(\lambda_{\text{cut}} - |\lambda|). \quad (7)$$

N_{tot} is the total dimension of the Dirac matrix. The recipe for LMA in terms of the CAA master Eq. (2) is shown in left column of Table I. Although LMA is particularly good for observables dominated by low-modes, such as the single pion state for lighter fermion masses, LMA does not work so well for heavier hadrons or when the quark mass is heavier [16, 18] (see also [22] for dependence on parity of states and (non-)Hermiticity of Dirac operators). This is due to the truncation of the sum in (6), *i.e.*, $f_{\text{LM}}(\lambda) = 0$ for $|\lambda| > \lambda_{\text{cut}}$.

One could improve the above by constructing a polynomial for $1/\lambda$ and using it to obtain a better (all-mode) approximation of the propagator above λ_{cut} :

$$S_{\text{AM}}(x, y) = \sum_{i=0}^{N_{\text{tot}}} v_i(x) f_{\text{AM}}(\lambda_i) v_i^\dagger(y), \quad (8)$$

$$f_{\text{AM}}(\lambda) = \begin{cases} \frac{1}{\lambda} & |\lambda| \leq \lambda_{\text{cut}} \\ P_n(\lambda) & |\lambda| > \lambda_{\text{cut}} \end{cases} \quad (9)$$

where $P_n(\lambda) \approx 1/\lambda$ is a polynomial of degree n . From (8) and (9), one computes the approximate propagator using $P_n(D_H)$ in the subspace orthogonal to the eigenvectors below λ_{cut} ,

$$S_{\text{AM}} = \sum_{i=1}^{N_{\text{eig}}} v_i \frac{1}{\lambda_i} v_i^\dagger + P_n(D_H) \left(1 - \sum_{i=1}^{N_{\text{eig}}} v_i v_i^\dagger\right), \quad (10)$$

with number of low-modes N_{eig} . In analogy to LMA, we refer to the above as all-mode averaging (AMA). A recipe similar to LMA is shown in the right column of Table I.

As emphasized in [9] approximate eigenvectors can be used in LMA (and AMA) to reduce the cost of this part of

the calculation. We have not done that as we find the cost of computing them exactly is not too burdensome and is partly recouped in the deflation of the Dirac operator.

Among many different ways [23–25] to obtain $P_n(\lambda)$, one of the easiest is to use the polynomial implicitly generated by an iterative linear solver such as conjugate gradient (CG). For example (8) can be implemented as a CG solution using the low-mode approximation applied to the source vector b (the coefficients of P_n depend on b) as the starting vector, $S_{\text{LM}}b$, which is nothing but a deflated CG with iteration number set to the degree of the polynomial, n . One can either fix n (number of iterations) or the CG residual vector stopping criterion. Either satisfies the covariance condition (**appx-3**). This particular construction of $P_n(D_H)$ is called the truncated solver method (TSM) [21]. The difference with AMA is that TSM is applied in [21] to a random source, and the unbiased result is guaranteed by stochasticity while AMA relies on covariance, so it does not need the random source.

In [18] low-modes are utilized with Z_3 noise to compute many-to-all hadron correlation functions for variance reduction. One may also choose $N_{\text{eig}} = 0$, $\lambda_{\text{cut}} = 0$ in (9), *i.e.* not to use eigenvectors at all. This may be effective for heavier quark masses, but for lighter quarks one needs a larger degree polynomial for an accurate approximation and $N_{\text{eig}} > 0$ is likely more cost-effective.

To ensure unbiasedness one should check, on a few configurations, the covariance of the particular implementation of the approximation $\mathcal{O}^{(\text{appx})}[U^g] = \mathcal{O}^{(\text{appx})}.g[U]$ by computing the approximation explicitly on a transformed gauge field to compare with the original gauge field to see that they are equivalent to numerical precision.

To compare the LMA and AMA methods, we use the 2+1 flavor Domain-Wall fermion (DWF) ensemble generated by the RBC/UKQCD collaboration [26] with lattice size $24^3 \times 64$, extra dimension size $L_s = 16$, and Iwasaki gauge action ($\beta = 2.13$, or $a^{-1} = 1.73$ GeV). The low-modes of the Hermitian DWF Dirac operator are obtained using a 4D-even-odd-preconditioned, shifted Lanczos algorithm [11] with accuracy $\|(D_H - \lambda_i)v_i\|/\|v_i\| < 10^{-12}$. The eigen-modes are used for LMA as in Eqs. (6) and (7), to deflate the CG, and to evaluate the low-mode parts of both \mathcal{O} and $\mathcal{O}^{(\text{appx})}$, and similarly for AMA as in Eqs. (8) and (9). In this paper we compute 180 low-modes for light quark mass $m = 0.01$ and 400 low-modes for $m = 0.005$.

We adopt translational symmetry on the lattice as G and take N_G propagator source locations, starting from the origin, separated by 12 lattice units in space and 16 in time, and the total set of translations numbers $N_G = 2^3 \times 4 = 32$. For AMA, the stopping condition of the “sloppy CG” for our approximation is $\|D_H x - b\|/\|b\| < 3 \times 10^{-3}$ while it is 10^{-8} in [2]. Note that when using an even-odd preconditioned Dirac operator, LMA and AMA guarantee unbiased estimators for translations by

TABLE II. Correlation function relative statistical error for $N_{\text{conf}} = 109$ (separated by 40 trajectories) and $N_G = 32$. Nucleon (N), pseudoscalar (PS), and vector (V) channels. $m = 0.005$. Gaussian smeared sink is used for the nucleon, others are point sinks. Gaussian smeared source is used for all channels.

Hadron	t	Original [%]	LMA [%]	AMA [%]
N	4	6.9	5.0	1.5
	8	9.2	3.2	1.9
	12	23	4.8	3.5
PS	4	4.5	0.98	0.86
	12	4.9	0.91	0.86
	28	5.0	1.3	1.3
V	4	3.9	2.9	0.6
	8	5.2	2.1	1.1
	12	12	3.4	2.3

an even number of sites (**appx-3**). We have explicitly checked this in our calculations.

Table II lists the relative statistical errors for various hadronic two-point correlation functions computed using LMA, AMA, and the original CG method, for $m = 0.005$. All were obtained with the same Gaussian smeared sources and point (Gaussian) sinks for pseudoscalar and vector (Nucleon) used in [2]. At short distance ($t = 4$), there is no improvement between the original and LMA cases, except in the pseudoscalar (PS) channel. This is because the contribution of higher modes is still important in the short-distance region. Although for LMA N_G could be taken as large as the lattice size with modest cost, we set $N_G = 32$ since larger N_G is not effective due to correlations between nearby gauge fields in our examples. On the other hand, AMA dramatically reduces the errors (more than 4-6 \times) for all channels (and different momenta) and for all distances. In this example the variance reduction by AMA comes almost entirely from the second term in Eq. (3) since $r = \text{Corr}(\mathcal{O}, \mathcal{O}^{(\text{appx})})$ is very close to one ($r > 0.9999$ for $m = 0.005$), even though the residual stopping criterion used for $\mathcal{O}^{(\text{appx})}$ is loose (3×10^{-3}). For LMA at short distance $r \simeq 0.9$ so the error from $\mathcal{O}^{(\text{rest})}$ is significant. We also confirm that for the PS channel both LMA and AMA yield improvement, with $r > 0.997$ even in the short distance region, as suggested previously for LMA using overlap fermions [17, 18]. For $m = 0.01$ r is somewhat smaller ($r > 0.99$), so the contribution from $\mathcal{O}^{(\text{rest})}$ is more significant. Only 180 low-modes were used for $m = 0.01$.

Figure 1 shows the nucleon effective mass using LMA and AMA for the data in Tab. II, and Tab. III compares these to an earlier high statistics study of nucleon structure functions [2]. The right-most panel in Fig. 1 shows significant improvement of the effective mass plateau for

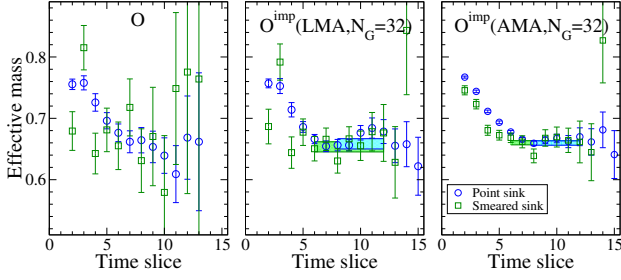


FIG. 1. Nucleon effective mass using LMA (middle) and AMA (right). $m = 0.005$. Unimproved calculation (left). See Tab. II for parameters. Colored bands denote fit mass and range. Gaussian sink.

TABLE III. Nucleon masses (GeV) using LMA, AMA and from data from a high statistics study [2]. See Table IV for costs. Gaussian (gauss) and point (pt) sinks.

			$\mathcal{O}^{(\text{imp})}, N_G = 32$		\mathcal{O}
m	sink	fit range	LMA	AMA	High stat.
0.005	pt	8-12	1.1391(145)	1.1413(61)	1.1561(104)
0.005	gauss	6-12	1.1305(143)	1.1420(58)	1.1481(100)
0.01	pt	9-15	1.2446(164)	1.2363(59)	1.2101(89)
0.01	gauss	7-15	1.2240(148)	1.2268(60)	1.2169(93)

AMA. Using the same fitting range, the precision of the nucleon mass attained with AMA is smaller by more than a factor of 1.5 compared to the high statistics study [2] where 3728 and 1424 measurements were made for $m = 0.005$ and 0.01 , respectively. The improved statistics make it easier to choose the fit range based on χ^2 , as seen in Fig. 1. LMA for nucleon masses was examined in [16].

Most of the cost of AMA comes from the low-mode and sloppy CG parts of the approximation $\mathcal{O}^{(\text{appx})}$ (deflating the Dirac operator significantly reduces the cost of computing $\mathcal{O}^{(\text{rest})}$), and the larger N_G , the lesser the relative cost of the former. The various costs for AMA in our examples are broken down in Table IV and compared to the high statistics study [2]. In the example using Gaussian sinks, AMA is roughly 16 and 5 times less expensive for roughly the same statistical error, for $m = 0.005$ and 0.01 , respectively. LMA is significantly less effective, 3.6 and 2.3 times less expensive. As N_G increases, AMA improves statistics with relatively little extra cost. For instance, for $N_G = 64$ AMA costs an additional 114, in units of the original propagator. The advantage of AMA clearly grows with increasing lattice size and decreasing quark mass. The cost of calculating

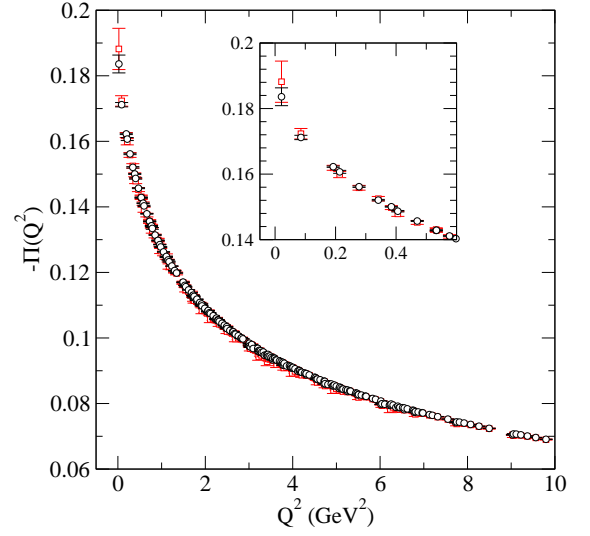


FIG. 2. Hadronic vacuum polarization from [1] (squares) and using AMA (circles). AMA achieves the same statistical error as the original calculation in the range 0-1 GeV^2 for about 2.6-20 times less computer time. See Tab. IV for details.

the correlation functions in this example is negligible, but this may not be the case for more complicated observables. Although disk space and CPU time for eigenvector I/O can be non-negligible, we ignore these as the costs strongly depend on the implementation details (*e.g.*, we could (de)compress eigenvectors) and the features of the I/O systems used.

Another impressive example of AMA is shown in Fig. 2, which depicts the hadronic vacuum polarization (HVP) from [1] and using AMA for roughly the same amount of computational resource (20 configurations, 1400 low-modes with accuracy $\|(D_H - \lambda_i)v_i\|/\|v_i\| < 10^{-10}$, $N_G = 708$, and sloppy CG stopping residual criterion 10^{-4} compared to 10^{-8} in [1]). The pion mass is $m_\pi = 476$ MeV and lattice size $48^3 \times 144$. The HVP contribution to the muon's anomalous magnetic moment is sensitive to the low Q^2 region [1], so constraining the HVP in this region is crucial to precisely extract the anomaly. In this test case (which was not optimized), to achieve the same errors on the HVP in the range 0-1 GeV^2 as the original calculation required about 2.6-20 times less computer time. Interestingly, LMA actually increases the error in this case by about $2-3\times$ because the low-modes do not saturate the Ward-Takahashi identity. The stopping criterion for $\mathcal{O}^{(\text{appx})}$ can not be too low for the same reason, though our choice may have been too conservative. The costs are summarized in Tab. IV. We note that in this case the cost of constructing the low mode part of the propagator is roughly equivalent to the sloppy CG cost, and that here again the contraction costs are negligible.

In this letter a new class of unbiased error reduction techniques is introduced, using approximations that are

TABLE IV. Computational cost. The unit of cost is one quark propagator without deflated CG, per configuration. $N_G = 32$ for nucleon masses and 708 for HVP. The last column gives the cost to achieve the same error for each method, normalized to [2] (nucleon mass m_N) and [1] (HVP) and scaled by the errors in Tab. III. HVP scaled costs are maximum and minimum in the range $Q^2 = 0 - 1$ GeV². For $m = 0.005$, in [2], non-relativistic spinors were used which means the scaled costs in this case were increased by two. The cost of $\mathcal{O}_G^{(\text{appx})}$ for AMA is split to show the sloppy CG and low-mode costs separately.

	N_{conf}	N_{meas}	LM	\mathcal{O}	$\mathcal{O}_G^{(\text{appx})}$	Tot.	scaled cost	
m_N	$m = 0.005, 400$ LM						gauss	pt
AMA	110	1	213	18	91+23	350	0.063	0.065
LMA	110	1	213	18	23	254	0.279	0.265
Ref. [2]	932	4	-	3728	-	3728 ^a	1	1
$m = 0.01, 180$ LM								
AMA	158	1	297	74	300+22	693	0.203	0.214
LMA	158	1	297	74	22	393	0.699	0.937
Ref. [2]	356	4	-	1424	-	1424	1	1
HVP	$m = 0.0036, 1400$ LM						max	min
AMA	20	1	96	11	504+420	1031	0.387	0.050
LMA	20	1	96	11	420	527	10.3	3.56
Ref. [1]	292	2	-	584	-	584	1	1

^a In [2] a doubled source was used to reduce this cost by two.

covariant under lattice symmetries. This is a generalization of low-mode averaging which reduces the statistical error for observables that are not dominated by low-modes. We have shown through several numerical examples that all-mode averaging is a powerful example of CAA, performing better than LMA and works well even in cases where LMA fails. In the examples given here, AMA reduced the cost by factors up to ~ 20 , over conventional computations, and these factors will only increase for larger lattice sizes and smaller quark masses. The method has great potential for investigations of difficult but important physics problems where statistical fluctuations still dominate the total uncertainty, like the nucleon electric dipole moment or hadronic contributions to the muon anomalous magnetic moment. Since CAA works without introducing any statistical bias (so long as condition **appx-3** holds), there are many possibilities that also satisfy **appx-1** and **appx-2**: One can construct $\mathcal{O}^{(\text{appx})}$ using different lattice fermions and parameters (mass, L_s (for DWF), boundary conditions and so on). $\langle \mathcal{O}_G^{(\text{appx})} \rangle$ can be measured on a larger number of gauge configurations, which is potentially advantageous for observables dominated by gauge noise such as disconnected diagrams. One may also consider other types of approximations such as the hopping parameter expansion used in [21], or approximations at the level of hadronic Green's functions.

Numerical calculations were performed using the RICC

at RIKEN and the Ds cluster at FNAL. We thank Sinya Aoki, Rudy Arthur, Gunnar Bali, Peter Boyle, Norman Christ, Thomas DeGrand, Leonardo Giusti, Shoji Hashimoto, Tomomi Ishikawa, Chulwoo Jung, Takashi Kaneko, Christoph Lehner, Meifeng Lin, Stefan Schaefer, Ruth Van de Water, Oliver Witzel, Takeshi Yamazaki, Jianglei Yu and other members of JLQCD, RBC, UKQCD collaborations for valuable discussions and comments. CPS QCD library [27] and other softwares (QMP, QIO) are used, which are supported by USQCD and US-DOE SciDAC program. This work was supported by the Japanese Ministry of Education Grant-in-Aid, Nos. 22540301 (TI), 23105714 (ES), 23105715 (TI) and U.S. DOE grants DE-AC02-98CH10886 (TI) and DE-FG02-92ER40716 (TB). We also thank BNL, the RIKEN BNL Research Center, and USQCD for providing resources necessary for completion of this work.

-
- [1] C. Aubin, T. Blum, M. Golterman, and S. Peris, (2012), arXiv:1205.3695 [hep-lat].
 - [2] T. Yamazaki, Y. Aoki, T. Blum, H.-W. Lin, S. Ohta, *et al.*, Phys.Rev. **D79**, 114505 (2009), arXiv:0904.2039 [hep-lat].
 - [3] E. Shintani, S. Aoki, N. Ishizuka, K. Kanaya, Y. Kikukawa, *et al.*, Phys.Rev. **D72**, 014504 (2005), arXiv:hep-lat/0505022 [hep-lat].
 - [4] F. Berruto, T. Blum, K. Orginos, and A. Soni, Phys. Rev. **D73**, 054509 (2006), arXiv:hep-lat/0512004.
 - [5] E. Shintani, S. Aoki, N. Ishizuka, K. Kanaya, Y. Kikukawa, *et al.*, Phys.Rev. **D75**, 034507 (2007), arXiv:hep-lat/0611032 [hep-lat].
 - [6] E. Shintani, S. Aoki, and Y. Kuramashi, Phys. Rev. **D78**, 014503 (2008), arXiv:0803.0797 [hep-lat].
 - [7] T. Blum, P. Boyle, N. Christ, N. Garron, E. Goode, *et al.*, Phys.Rev.Lett. **108**, 141601 (2012), arXiv:1111.1699 [hep-lat].
 - [8] M. J. Savage, Prog. Part. Nucl. Phys. **67**, 140 (2012), arXiv:1110.5943 [nucl-th].
 - [9] L. Giusti, P. Hernandez, M. Laine, P. Weisz, and H. Wittig, JHEP **04**, 013 (2004), arXiv:hep-lat/0402002.
 - [10] T. A. DeGrand and S. Schaefer, Comput. Phys. Commun. **159**, 185 (2004), arXiv:hep-lat/0401011.
 - [11] H. Neff, N. Eicker, T. Lippert, J. W. Negele, and K. Schilling, Phys. Rev. **D64**, 114509 (2001), arXiv:hep-lat/0106016.
 - [12] L. Giusti, C. Hoelbling, M. Luscher, and H. Wittig, Comput. Phys. Commun. **153**, 31 (2003), arXiv:hep-lat/0212012.
 - [13] T. A. DeGrand and S. Schaefer, Phys. Rev. **D72**, 054503 (2005), arXiv:hep-lat/0506021.
 - [14] M. Luscher, JHEP **07**, 081 (2007), arXiv:0706.2298 [hep-lat].
 - [15] H. Fukaya *et al.* (JLQCD), Phys. Rev. Lett. **98**, 172001 (2007), arXiv:hep-lat/0702003.
 - [16] L. Giusti and S. Necco, PoS **LAT2005**, 132 (2006), arXiv:hep-lat/0510011.
 - [17] J. Noaki *et al.* (JLQCD and TWQCD), Phys. Rev. Lett. **101**, 202004 (2008), arXiv:0806.0894 [hep-lat].

- [18] A. Li *et al.* (xQCD), Phys. Rev. **D82**, 114501 (2010), arXiv:1005.5424 [hep-lat].
- [19] G. S. Bali, H. Neff, T. Duessel, T. Lippert, and K. Schilling (SESAM), Phys. Rev. **D71**, 114513 (2005), arXiv:hep-lat/0505012.
- [20] J. Foley, K. Jimmy Juge, A. O’Cais, M. Peardon, S. M. Ryan, *et al.*, Comput.Phys.Commun. **172**, 145 (2005), arXiv:hep-lat/0505023 [hep-lat].
- [21] G. S. Bali, S. Collins, and A. Schafer, Comput. Phys. Commun. **181**, 1570 (2010), arXiv:0910.3970 [hep-lat].
- [22] G. Bali, L. Castagnini, and S. Collins, PoS **LAT-TICE2010**, 096 (2010), arXiv:1011.1353 [hep-lat].
- [23] M. Luscher, Nucl.Phys. **B418**, 637 (1994), arXiv:hep-lat/9311007 [hep-lat].
- [24] A. Borici and P. de Forcrand, Nucl.Phys. **B454**, 645 (1995), arXiv:hep-lat/9505021 [hep-lat].
- [25] W. Kamleh and M. Peardon, Comput.Phys.Commun. **183**, 1993 (2012), arXiv:1106.5625 [hep-lat].
- [26] Y. Aoki *et al.* (RBC), Phys. Rev. **D83**, 074508 (2011), arXiv:1011.0892 [hep-lat].
- [27] Columbia Physics System code developed and maintained by members of RBC/UKQCD collaborations <http://qcdoc.phys.columbia.edu/cps.html>, .